CRUK AND BIG DATA

DR FIONA REDDINGTON
HEAD OF POPULATION RESEARCH FUNDING

Cancer Research UK

WHAT WE SPEND

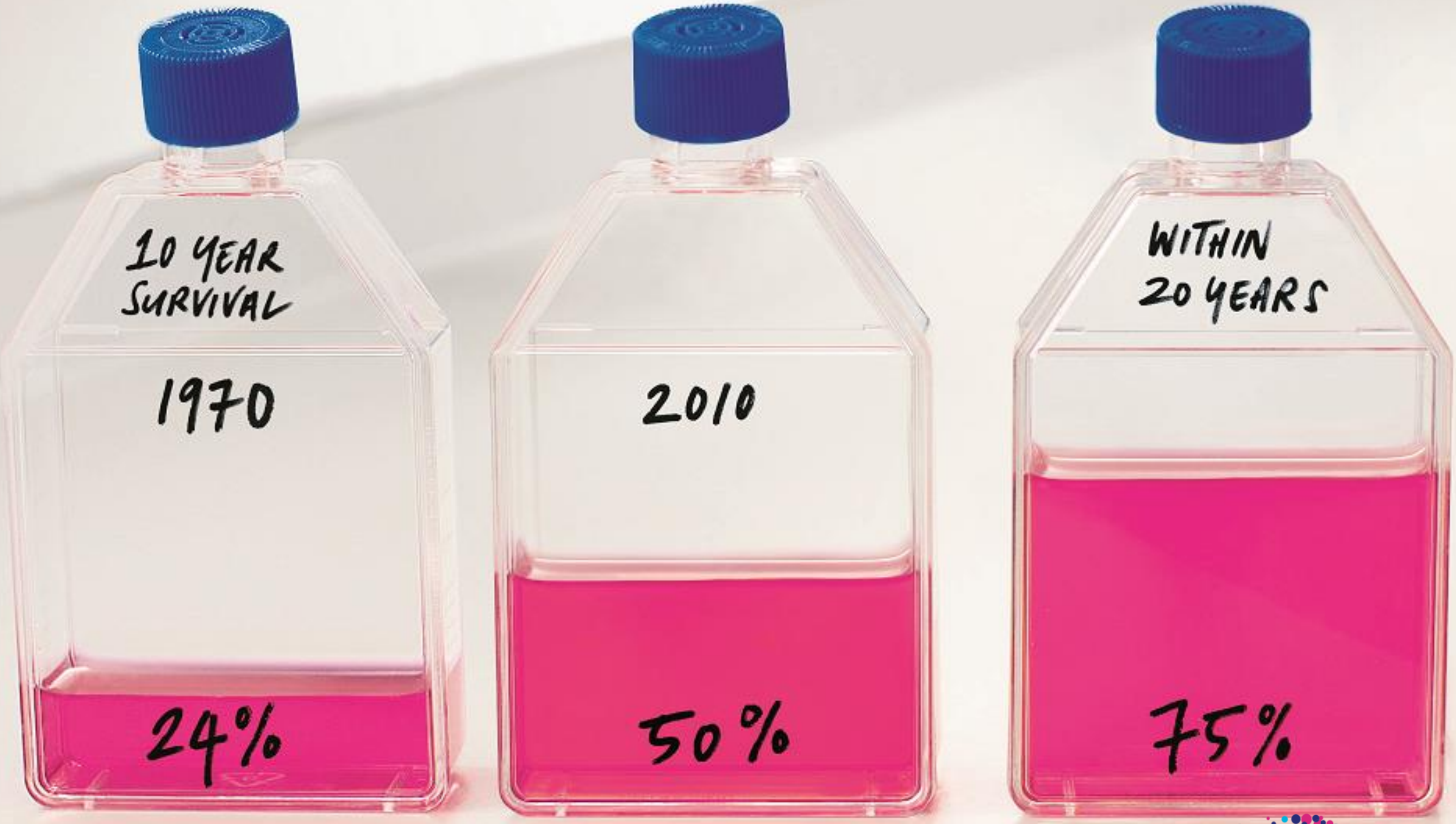
- We spent £341m on research in 2014/15
- The money we raise is spent on
  - research
  - information
  - advocacy and public policy

WHERE WE WORK

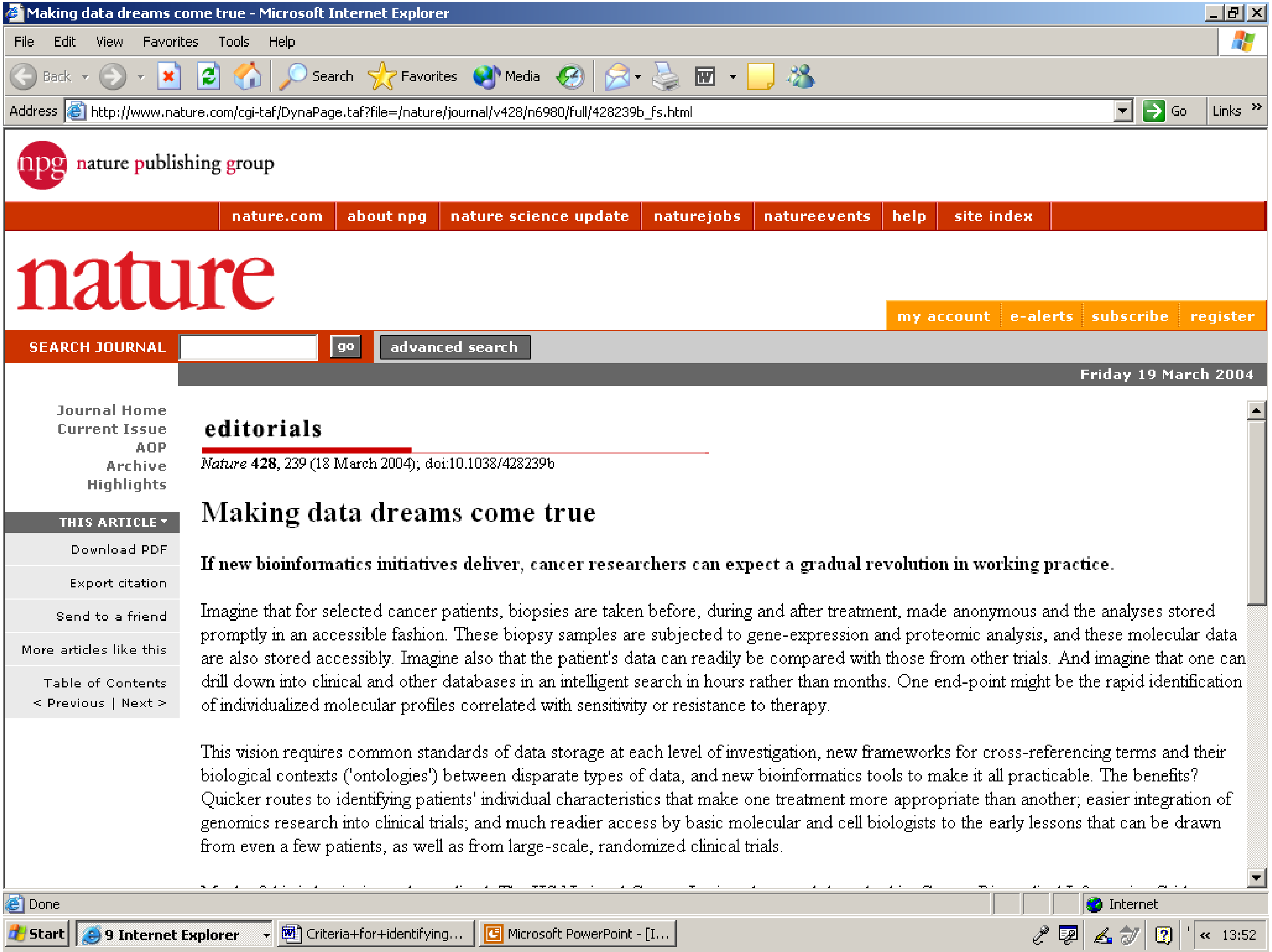
- Cancer Research UK supports over 500 research groups
- We support research in about 40 towns and cities across the UK

CANCER
RESEARCH
UK

Cancer Research UK's Ambition



CANCER
RESEARCH
UK


Making data dreams come true - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v428/n6980/full/428239b_fs.html
Go Links
nature publishing group
nature.com about npg nature science update naturejobs natureevents help site index
nature
my account e-alerts subscribe register
SEARCH JOURNAL go advanced search
Friday 19 March 2004
Journal Home
Current Issue
AOP
Archive
Highlights
THIS ARTICLE
Download PDF
Export citation
Send to a friend
More articles like this
Table of Contents
< Previous | Next >
Done
Internet
Start 9 Internet Explorer Criteria+for+identifying... Microsoft PowerPoint - [I...
13:52

editorials

Nature **428**, 239 (18 March 2004); doi:10.1038/428239b

Making data dreams come true

If new bioinformatics initiatives deliver, cancer researchers can expect a gradual revolution in working practice.

Imagine that for selected cancer patients, biopsies are taken before, during and after treatment, made anonymous and the analyses stored promptly in an accessible fashion. These biopsy samples are subjected to gene-expression and proteomic analysis, and these molecular data are also stored accessibly. Imagine also that the patient's data can readily be compared with those from other trials. And imagine that one can drill down into clinical and other databases in an intelligent search in hours rather than months. One end-point might be the rapid identification of individualized molecular profiles correlated with sensitivity or resistance to therapy.

This vision requires common standards of data storage at each level of investigation, new frameworks for cross-referencing terms and their biological contexts ('ontologies') between disparate types of data, and new bioinformatics tools to make it all practicable. The benefits? Quicker routes to identifying patients' individual characteristics that make one treatment more appropriate than another; easier integration of genomics research into clinical trials; and much readier access by basic molecular and cell biologists to the early lessons that can be drawn from even a few patients, as well as from large-scale, randomized clinical trials.

What do our researchers need to be able to do with data?

Discover	Acquire	Store	Share	Mine
🟧	🟩	🟧	🟧	🟧

- Collection of standardised datasets is becoming more commonplace for both research and clinical data
- Data sharing needs to improve
- Move towards provision of large scale datasets and infrastructure - not every research project needs to collect every data item
- Identify where re-use is feasible and where do we still need to fund collection of specific data items?

Generating Big Data is expensive

- PROSPECTIVE COHORT OF 500,000 PARTICIPANTS
- LIFESTYLE DATA, MEASUREMENTS AND BIOLOGICAL SAMPLES COLLECTED AT BASELINE

COST = £90 MILLION FOR 5 YEARS

- FOR £20 PER PARTICIPANT CAN COLLECT AND ANALYSE APPROX 50 BIOMARKERS – WILL TAKE 18 MONTHS

COST = £10 MILLION TO GENERATE THE DATASET

- CURRENT BID TO ADD IMAGING ENHANCEMENTS ON SUBSET OF 100,000 PEOPLE (MRI, DEXA, 3D ULTRASOUND)

COST = £6 MILLION FOR PILOT, £26 MILLION FOR 5 YEAR STUDY

- PLAN TO TRACK DISEASE OUTCOMES VIA LINKAGE TO ROUTINE COLLECTED MEDICAL DATA (E.G. HES, GPES, REGISTRY DATA)

COST = ? TBC WITH DATA PROVIDERS

THIS IS THE COST OF THE ESTABLISHMENT OF THE RESOURCE AND DATA GENERATION – IT DOES **NOT** INCLUDE THE COSTS OF ANY ASSOCIATED RESEARCH PERFORMED ON THE DATA

Why is Big Data important in cancer research?

Discovery	Clinical	Population
ICGC	YODA	UK Biobank
Human Genome project	CPRD	Cohort studies
Actionable Genome Consortium	CSDR.com	Routinely collected datasets (NHS and beyond)
	NPG Data Disclosure Project	
100,000 genome, CRUK Strat Med		
GA4GH		

And it's not just the Big Data…….

CANCER
RESEARCH
UK

An example from clinical trials...

- Heterogeneity
  - Not all applications will be run through a CRUK Trials Unit (CTU)
  - 8 CRUK CTU's – all free to choose the data management software/systems they use
  - Multiple accreditation bodies (ECRIN, UKCRC, etc)
  - CTU's do not just work on cancer trials
  - CTU's do not just work on UK trials

Any guideline/policy/recommendation needs to be internationally compatible, system independent and non disease specific

CANCER
RESEARCH
UK

Future trials - What do we need?

- Data providers
  - Encourage applicants to think about entire data life cycle upfront when designing trial/study
    - Trial/cohort registration
    - Data discoverability – what data/what format?
    - Data accessibility – process for requesting data, how are decisions made, how many requests y/n, etc
  - Work to existing standards where they exist (e.g. CDISC, HL7, CONSORT, STROBE, etc)
  - Need common framework to work to where standards are not yet defined

CANCER RESEARCH UK

Future trials - What do we need?

- Data requestors
  - What level of data is actually needed?
    - Raw data, CRF's, lay summaries
  - Is the data being requested for an appropriate purpose?
    - statistical analysis plan
    - valid scientific question
    - appropriate level of data requested
  - How will the results be made available and original team appropriately credited?

CANCER RESEARCH UK

What about trials that are closed/already underway?

- Need a rational way to prioritise:
  - What studies this should apply to
  - How far back should this apply?
  - How to resource – especially if funding for the trial has now ceased?
  - Does the data still exist!

CANCER
RESEARCH
UK

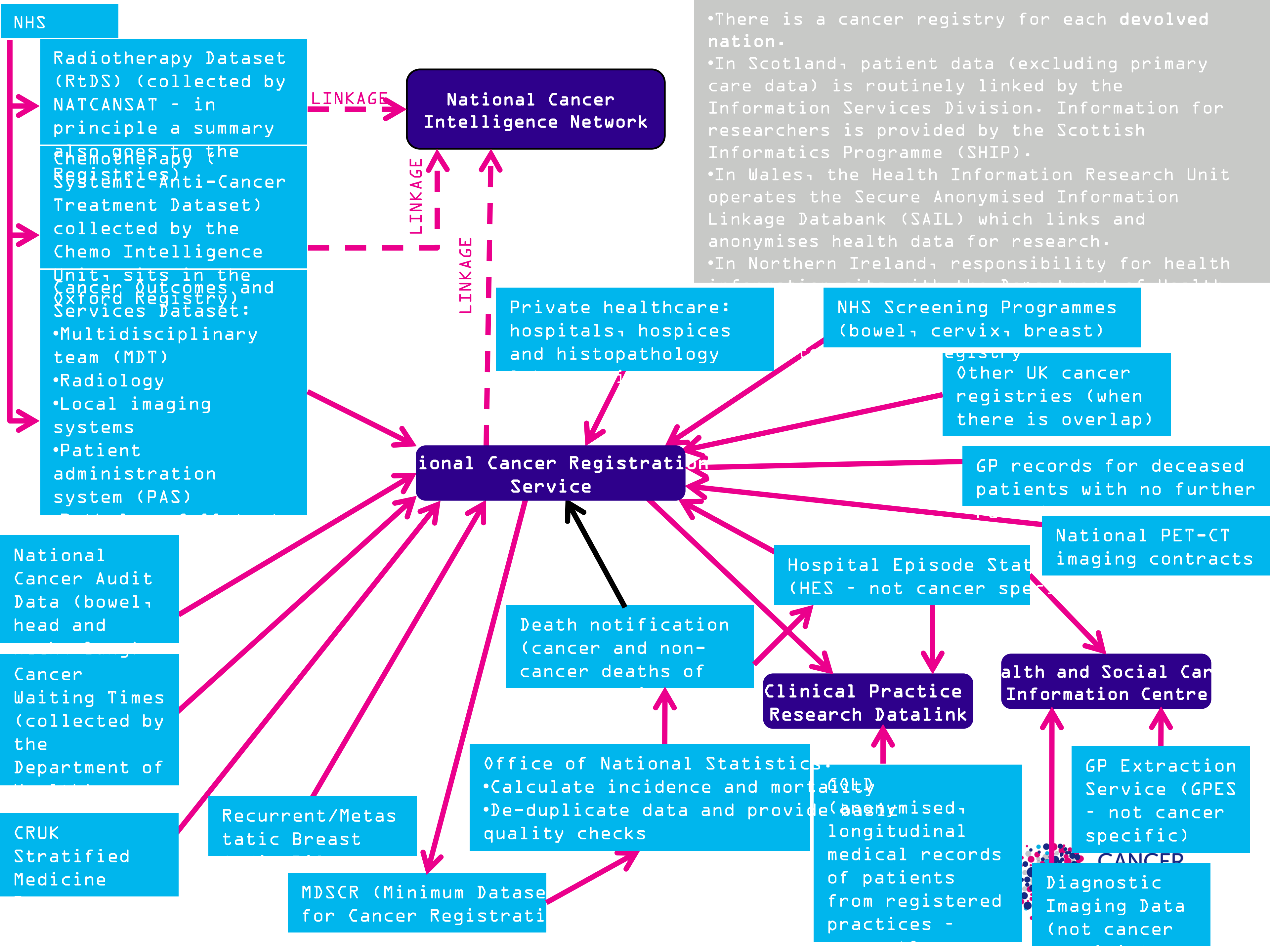
What about routinely collected datasets...

- Data have already been collected

- Data are discoverable

- Storage is taken care of by 3rd party provider

- So all our researchers need to do is access the data via agreed procedures/mechanisms



NHS

Radiotherapy Dataset (RtDS) (collected by NATCANSAT – in principle a summary also goes to the Registries)

Chemotherapy (Systemic Anti-Cancer Treatment Dataset) collected by the Chemo Intelligence Unit, sits in the Oxford Registry)

Cancer Outcomes and Services Dataset:

- Multidisciplinary team (MDT)
- Radiology
- Local imaging systems
- Patient administration system (PAS)

LINKAGE

LINKAGE

LINKAGE

National Cancer Intelligence Network

- There is a cancer registry for each **devolved nation**.
- In Scotland, patient data (excluding primary care data) is routinely linked by the Information Services Division. Information for researchers is provided by the Scottish Informatics Programme (SHIP).
- In Wales, the Health Information Research Unit operates the Secure Anonymised Information Linkage Databank (SAIL) which links and anonymises health data for research.
- In Northern Ireland, responsibility for health

ional Cancer Registration Service

Private healthcare: hospitals, hospices and histopathology

NHS Screening Programmes (bowel, cervix, breast)

Other UK cancer registries (when there is overlap)

GP records for deceased patients with no further

National PET-CT imaging contracts

Hospital Episode Stat (HES – not cancer spe

National Cancer Audit Data (bowel, head and

Cancer Waiting Times (collected by the Department of

CRUK Stratified Medicine

Death notification (cancer and non-cancer deaths of

Clinical Practice Research Datalink

alth and Social Car Information Centre

Office of National Statistics:

- Calculate incidence and mortality
- De-duplicate data and provide basic quality checks

GOLD (anonymised, longitudinal medical records of patients from registered practices –

GP Extraction Service (GPES – not cancer specific)

Recurrent/Metastatic Breast

MDSCR (Minimum Dataset for Cancer Registrati

Diagnostic Imaging Data (not cancer

Informatics isn’t just about data….

- We are entering an era where data generation is exploding

- Shortage of skills to undertake complex data linkage and analysis
- Methodological research needs development
- Understanding the science!
- Linking to data from other domains

  (e.g. National Pupil Database, Individual Learner Records, UCAS application records, Higher Education Statistics Data, Work and Pensions Longitudinal Study)

CANCER RESEARCH UK

So what about preservation….

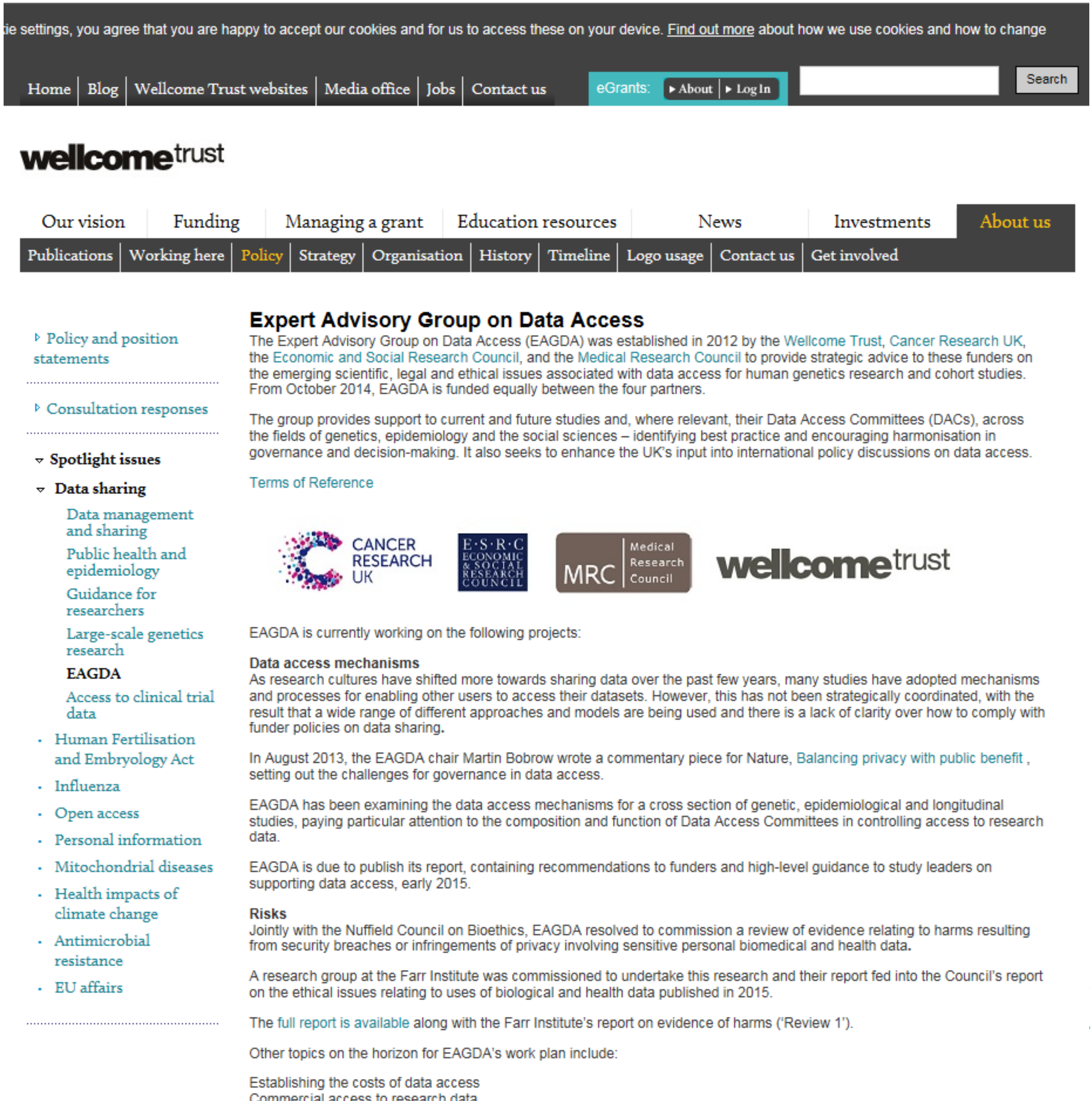
- Funder policies expect appropriate sharing, curation and preservation throughout data life cycle
  - responsibility of data custodians
  - does not always align with funding cycles

- Linking to data from other domains is important – other models to learn from?

- This requires partnership

CANCER
RESEARCH
UK

How can we partner effectively?



ie settings, you agree that you are happy to accept our cookies and for us to access these on your device. Find out more about how we use cookies and how to change

Expert Advisory Group on Data Access

The Expert Advisory Group on Data Access (EAGDA) was established in 2012 by the Wellcome Trust, Cancer Research UK, the Economic and Social Research Council, and the Medical Research Council to provide strategic advice to these funders on the emerging scientific, legal and ethical issues associated with data access for human genetics research and cohort studies. From October 2014, EAGDA is funded equally between the four partners.

The group provides support to current and future studies and, where relevant, their Data Access Committees (DACs), across the fields of genetics, epidemiology and the social sciences – identifying best practice and encouraging harmonisation in governance and decision-making. It also seeks to enhance the UK's input into international policy discussions on data access.

Terms of Reference

EAGDA is currently working on the following projects:

Data access mechanisms
As research cultures have shifted more towards sharing data over the past few years, many studies have adopted mechanisms and processes for enabling other users to access their datasets. However, this has not been strategically coordinated, with the result that a wide range of different approaches and models are being used and there is a lack of clarity over how to comply with funder policies on data sharing.

In August 2013, the EAGDA chair Martin Bobrow wrote a commentary piece for Nature, Balancing privacy with public benefit , setting out the challenges for governance in data access.

EAGDA has been examining the data access mechanisms for a cross section of genetic, epidemiological and longitudinal studies, paying particular attention to the composition and function of Data Access Committees in controlling access to research data.

EAGDA is due to publish its report, containing recommendations to funders and high-level guidance to study leaders on supporting data access, early 2015.

Risks
Jointly with the Nuffield Council on Bioethics, EAGDA resolved to commission a review of evidence relating to harms resulting from security breaches or infringements of privacy involving sensitive personal biomedical and health data.

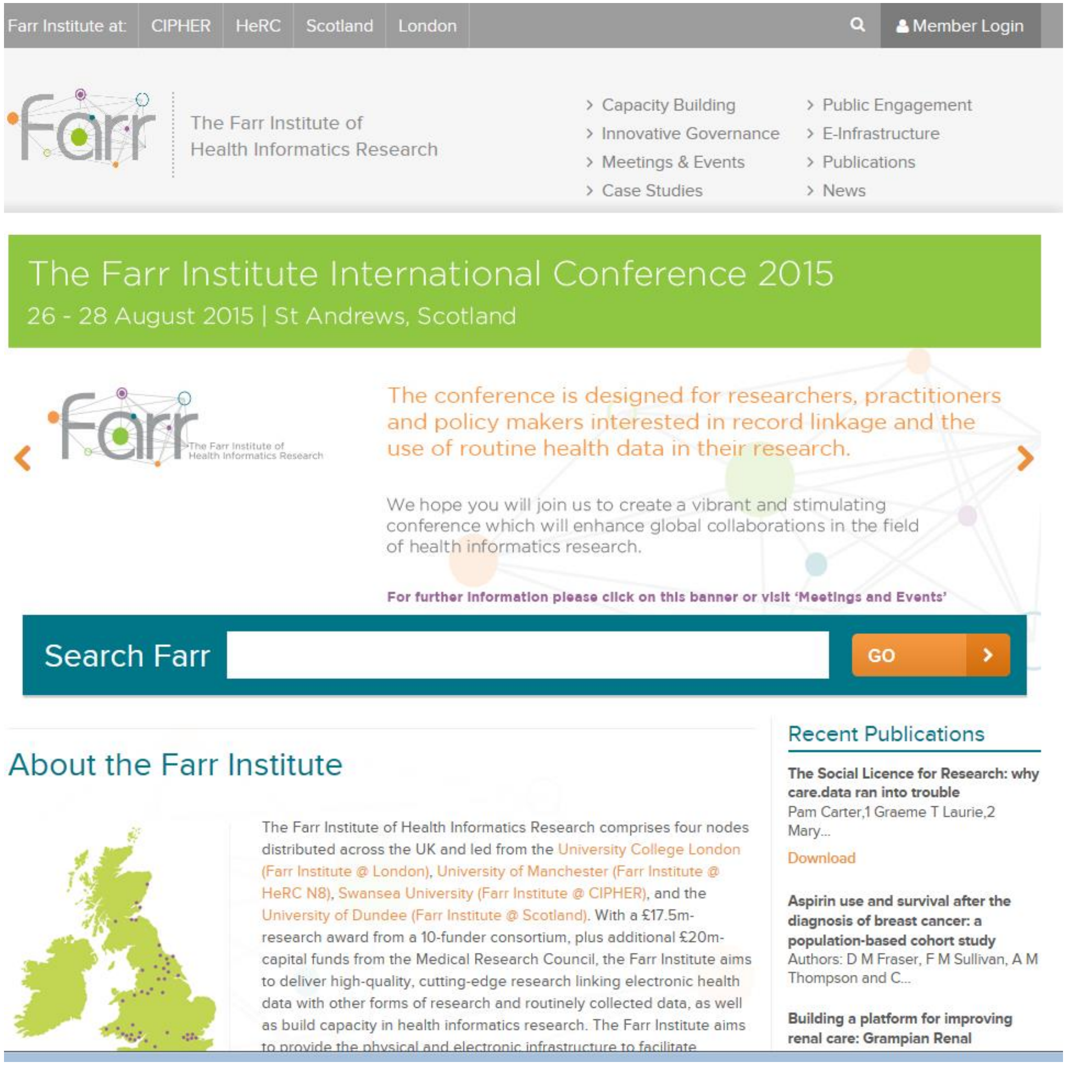
A research group at the Farr Institute was commissioned to undertake this research and their report fed into the Council's report on the ethical issues relating to uses of biological and health data published in 2015.

The full report is available along with the Farr Institute's report on evidence of harms ('Review 1').

Other topics on the horizon for EAGDA's work plan include:

Establishing the costs of data access
Commercial access to research data

CANCER RESEARCH UK



The Farr Institute International Conference 2015

26 - 28 August 2015 | St Andrews, Scotland

The conference is designed for researchers, practitioners and policy makers interested in record linkage and the use of routine health data in their research.

We hope you will join us to create a vibrant and stimulating conference which will enhance global collaborations in the field of health informatics research.

For further information please click on this banner or visit 'Meetings and Events'

About the Farr Institute

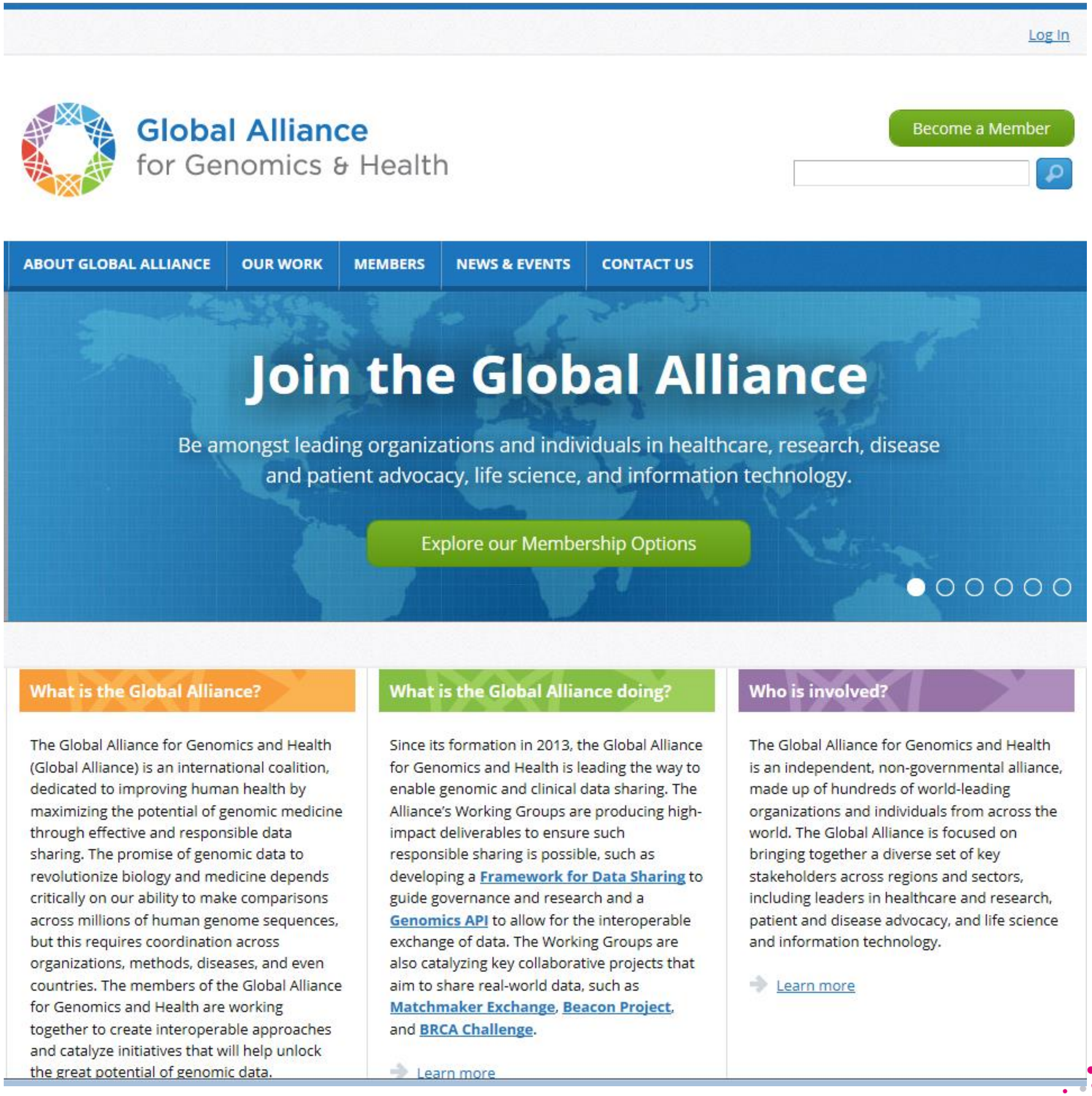
The Farr Institute of Health Informatics Research comprises four nodes distributed across the UK and led from the University College London (Farr Institute @ London), University of Manchester (Farr Institute @ HeRC N8), Swansea University (Farr Institute @ CIPHER), and the University of Dundee (Farr Institute @ Scotland). With a £17.5m-research award from a 10-funder consortium, plus additional £20m-capital funds from the Medical Research Council, the Farr Institute aims to deliver high-quality, cutting-edge research linking electronic health data with other forms of research and routinely collected data, as well as build capacity in health informatics research. The Farr Institute aims to provide the physical and electronic infrastructure to facilitate

Recent Publications

The Social Licence for Research: why care.data ran into trouble
Pam Carter,1 Graeme T Laurie,2 Mary...
Download

Aspirin use and survival after the diagnosis of breast cancer: a population-based cohort study
Authors: D M Fraser, F M Sullivan, A M Thompson and C...

Building a platform for improving renal care: Grampian Renal



Join the Global Alliance

Be amongst leading organizations and individuals in healthcare, research, disease and patient advocacy, life science, and information technology.

What is the Global Alliance?

The Global Alliance for Genomics and Health (Global Alliance) is an international coalition, dedicated to improving human health by maximizing the potential of genomic medicine through effective and responsible data sharing. The promise of genomic data to revolutionize biology and medicine depends critically on our ability to make comparisons across millions of human genome sequences, but this requires coordination across organizations, methods, diseases, and even countries. The members of the Global Alliance for Genomics and Health are working together to create interoperable approaches and catalyze initiatives that will help unlock the great potential of genomic data.

What is the Global Alliance doing?

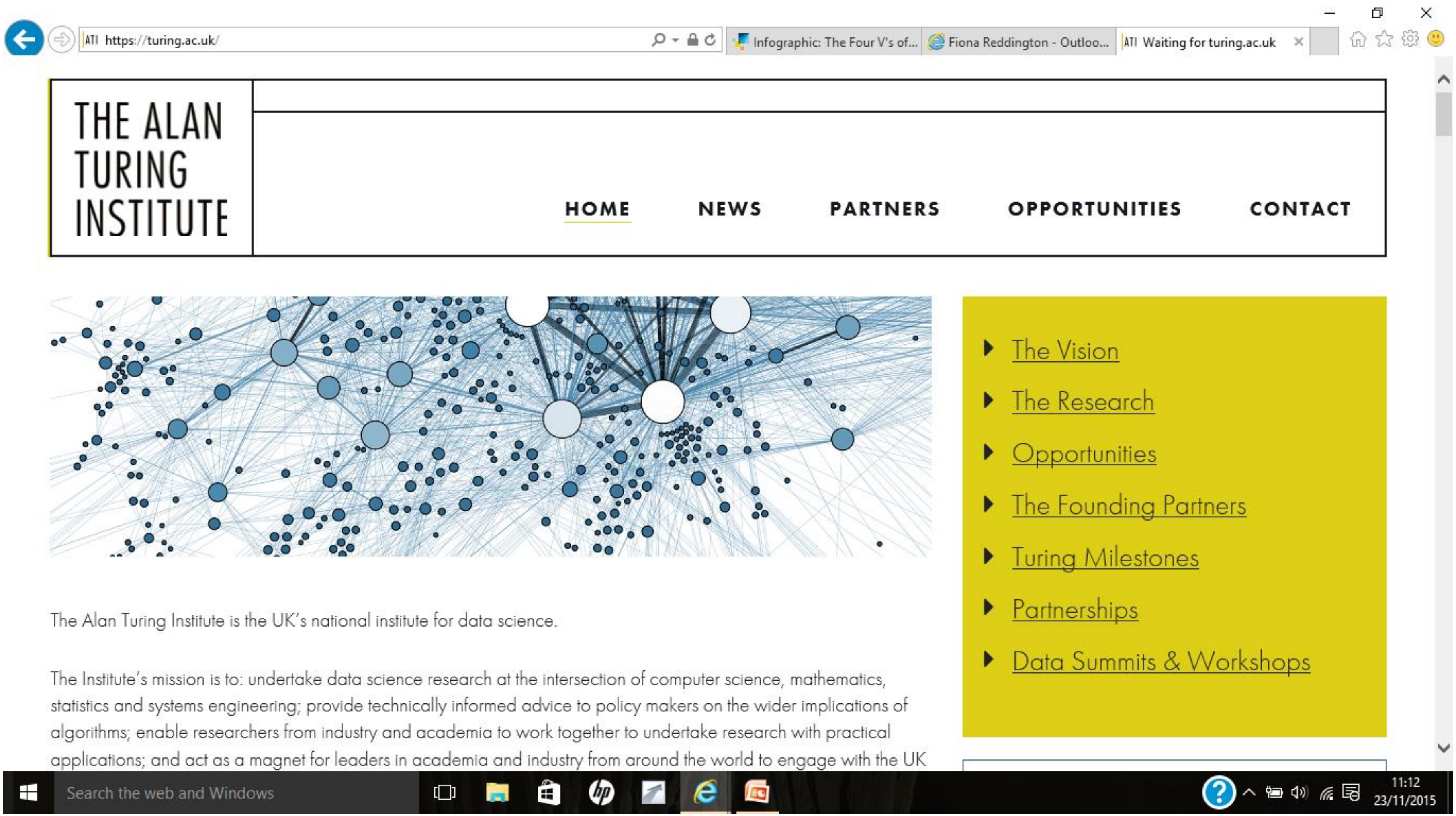
Since its formation in 2013, the Global Alliance for Genomics and Health is leading the way to enable genomic and clinical data sharing. The Alliance's Working Groups are producing high-impact deliverables to ensure such responsible sharing is possible, such as developing a **Framework for Data Sharing** to guide governance and research and a **Genomics API** to allow for the interoperable exchange of data. The Working Groups are also catalyzing key collaborative projects that aim to share real-world data, such as **Matchmaker Exchange**, **Beacon Project**, and **BRCA Challenge**.

Learn more

Who is involved?

The Global Alliance for Genomics and Health is an independent, non-governmental alliance, made up of hundreds of world-leading organizations and individuals from across the world. The Global Alliance is focused on bringing together a diverse set of key stakeholders across regions and sectors, including leaders in healthcare and research, patient and disease advocacy, and life science and information technology.

Learn more


https://turing.ac.uk/
Infographic: The Four V's of...
Fiona Reddington - Outloo...
Waiting for turing.ac.uk
Search the web and Windows
11:12
23/11/2015

THE ALAN TURING INSTITUTE

HOME NEWS PARTNERS OPPORTUNITIES CONTACT

- The Vision
- The Research
- Opportunities
- The Founding Partners
- Turing Milestones
- Partnerships
- Data Summits & Workshops

The Alan Turing Institute is the UK's national institute for data science.

The Institute's mission is to: undertake data science research at the intersection of computer science, mathematics, statistics and systems engineering; provide technically informed advice to policy makers on the wider implications of algorithms; enable researchers from industry and academia to work together to undertake research with practical applications; and act as a magnet for leaders in academia and industry from around the world to engage with the UK

THANK YOU!

cruk.org

Dr Fiona Reddington

Head of Population, Prevention and
Behavioural Research

020 3469 5324

fiona.reddington@cancer.org.uk

Angel Building, 407 St John Street
London EC1V 4AD

CANCER
RESEARCH
UK